PhD student: Stijn De Beugher
Promotor: Toon Goedemé
Co-promotor: Tinne Tuytelaars, Geert Brône

Computer vision techniques for automatic analysis of eye-tracking data

Research focus

- This PhD research focuses on a novel method for the semi-automatic analysis of mobile eye-tracking data in natural environments.
- Real life experiments out of lab conditions
- Data analysis of such a recording:

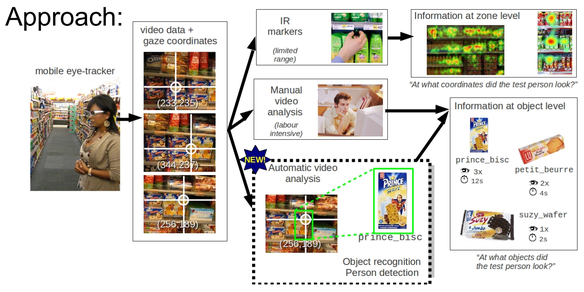
  "How often and for how long did a person look a specific object or person?"
- Data analysis is hard: moving camera & moving objects in scene
- Current methods:
  - Manual coding (Time-consuming, thus expensive)
  - Marker based analysis (Area Of Analysis needs to be defined before test, objects must stay fixed within plane)

Paradigm shift: object based instead of coordinate based analysis.

Goal

- Integration of image processing techniques
- Benefits:
  - Target of analysis is no longer restricted to a fixed region
  - Manual labour is limited
  - (Semi-)automatic detection of specific objects, bodies, faces, hands,…
  - Automatic generation of userfriendly output: statistical data, time line representation, object cloud, heat map, ELAN file.

- Approach:



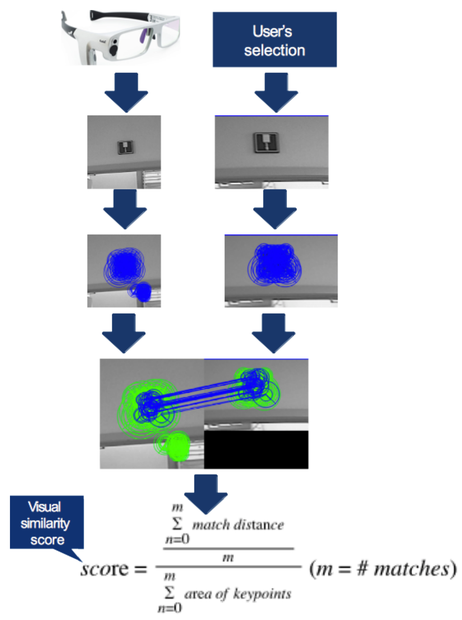
Applications:
- shopping research:

- human-human interaction:

Approach

Object recognition

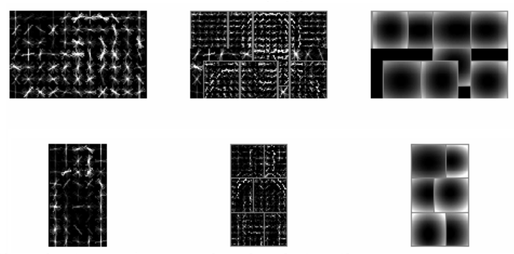
- Object recognition techniques to count how often a specific object was viewed.
- Application: shopping research
- Feature based detection: ORB[3] features
- Calculation of visual similarity score



$$score = \frac{\frac{\sum_{n=0}^{m} match\ distance}{m}}{\frac{m}{\sum_{n=0} area\ of\ keypoints}} \quad (m = \#\ matches)$$

Person detection

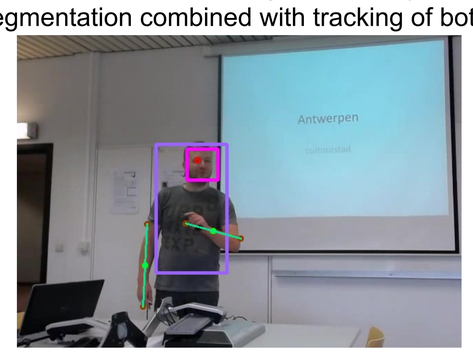
- Part based person detection (DPM) to count how often one looked at another person / face.
- Application: human-human interaction, customer journey analysis
- Own trained human torso model (upper 60% of PASCAL VOC[2] model)



- Classification in torso/upper body detection
- Distinguish between looking to someone from large distance (torso) and standing in front of someone (upper body)
- Implementation of a temporal smoothing technique:
  - Use gaze-date to improve detection rate
  - Assume that a visual fixation should take at least 150ms

Hand detection

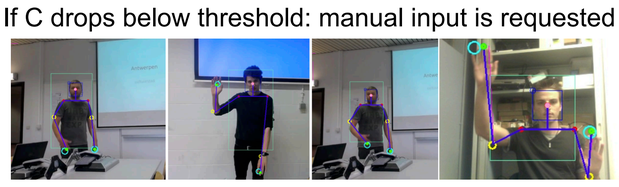
- Skin based hand detection to count how often one looked at the hands of another person
- Application: human-human interaction, gesture analysis,…
- Accurate skin segmentation combined with tracking of both hand, wrist and elbow.



- Integration of a semi-automatic approach:
  - Confidence of each hand candidate is validated against probability maps:

    $C = \{(D>D_{max}) \wedge (pred>pred_{max})\} \vee \{max(P_{Elbow},P_{Wrist}) < P_{TH}\}$

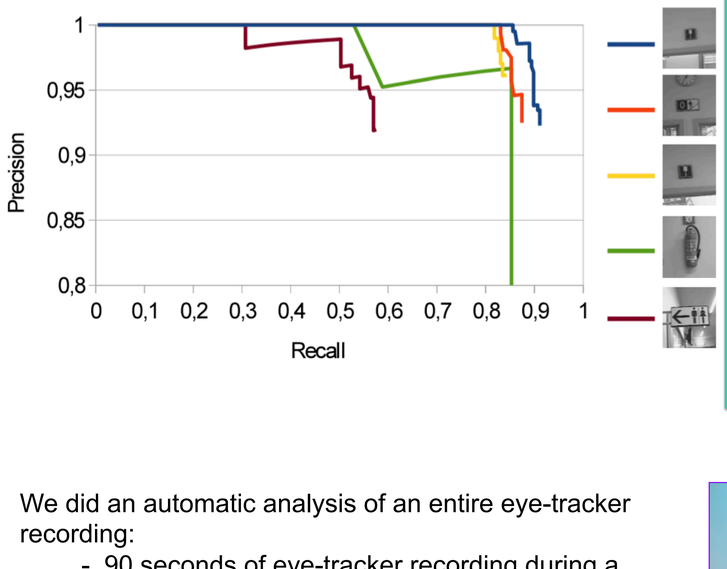
  - If C drops below threshold: manual input is requested



Results

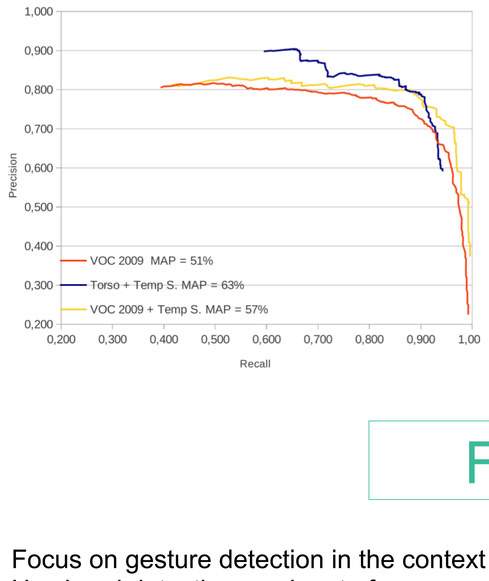
Object recognition

- Highly accurate object recognition
- Validation was done on a dataset of 2000 labeled frames
- Processing takes as long as video itself when up to 5 objects of interest are selected



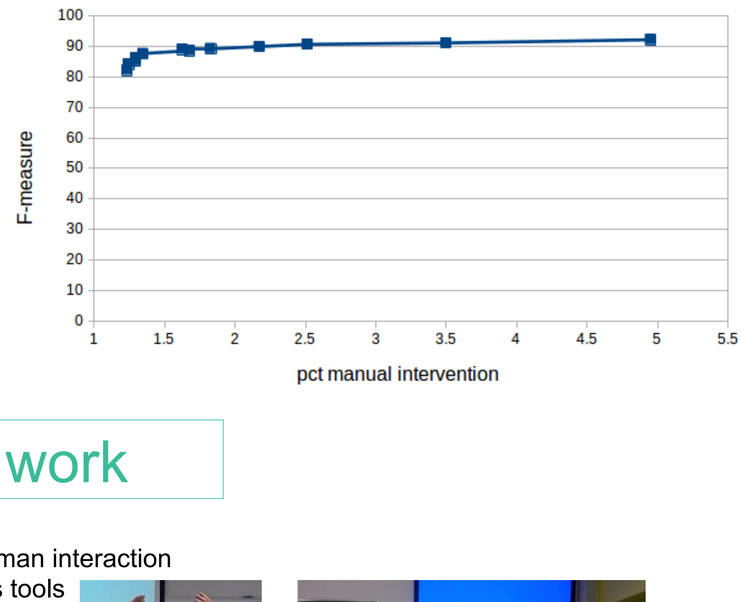
Person detection

- Highly accurate torso and upper body detection
- Gaze based temporal smoothing achieves high accuracy
- Validation was done using a recorded dataset of 3000 frames
- Low computational cost due to optimization using FFLD[1] approach



Hand detection

- Highly accurate hand detection, even with a low amount of manual intervention
- Validation was done on a dataset containing 6000 hand labels
- The amount of manual intervention

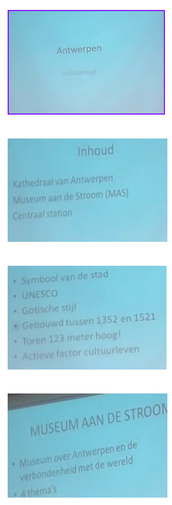


Future work

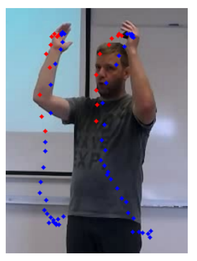
- We did an automatic analysis of an entire eye-tracker recording:
  - 90 seconds of eye-tracker recording during a presentation
  - Objects of interest: 4 slides and torso / upper body
  - Comparison with manual analysis:

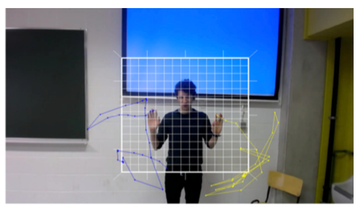
Automatic vs manual	Level
Agreement	90,4%
Scott's Pi	85,7%
Cohen's Kappa	85,8%
Krippendorff's Alpha	85,7%

  - Autmatic analysis took only 5 minutes, manual labelling took 20 minutes



Focus on gesture detection in the context of human-human interaction
Use hand detections as input of more compext analysis tools
Detection of basic gestures:
- gestural cueing
- pointing
- beats / batons





[1] C. Dubout and F. Fleuret. Exact Acceleration of Linear Object Detectors. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2012.
[2] The PASCAL Visual Object Classes Challenge 2009 (VOC2009) Dataset http://www.pascal-network.org/challenges/VOC/voc2009/workshop/index.html
[3] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: An efficient alternative to sift or surf. In ICCV, pages 2564–2571.